

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

DESCRIPTION

CLK PROTEIN KINASES AND RELATED PRODUCTS AND METHODS

RELATED APPLICATION

This application claims the benefit of PCT Application WO 97/48723 (PCT/IB97/00946), published on December 24, 1997, filed on June 17, 1997, Axel Ullrich et al., entitled "Novel PTP20, PCP-2, BDP-1, CLK, and SIRP Proteins and Related Products and Methods," and United States Provisional Patent Application Serial No. 60/034,286, entitled "CLK Protein Kinases and Related Products and Methods," Axel Ullrich and Oliver Nayler, filed December 19, 1996, both of which are incorporated by reference herein in their entirety including all figures, tables, and drawings.

INTRODUCTION

The present invention relates to novel CDC2 like protein kinases (CLK protein kinases). These protein kinases phosphorylate proteins rich in serine and arginine.

BACKGROUND OF THE INVENTION

The following description of the background of the invention is provided to aid in understanding the invention, but is not admitted to be or describe prior art to the invention.

Cellular signal transduction is a fundamental mechanism whereby extracellular stimuli are relayed to the interior of cells and subsequently regulate diverse cellular processes. One of the key biochemical mechanisms of signal transduction involves the reversible phosphorylation of proteins. Phosphorylation of polypeptides regulates the activity of mature proteins by altering their structure and function. Phosphate most often resides on the hydroxyl moiety (-OH) of serine, threonine, or tyrosine amino acids in proteins. Enzymes that mediate phosphorylation of cellular effectors fall into two classes. While protein phosphatases hydrolyze phosphate moieties from phosphoryl protein substrates, protein kinases transfer a phosphate moiety from adenosine triphosphate to protein substrates. The converse functions of protein kinases and protein phosphatases balance and regulate the flow of signals in signal transduction processes.

Protein kinases and protein phosphatases are typically divided into two groups: receptor and non-receptor type proteins. Receptor protein kinases are comprised of an extracellular domain, a membrane spanning region, and a catalytic domain.

Protein kinases and protein phosphatases are divided further into three classes based upon the amino acids they act upon. Some catalyze the addition or hydrolysis of phosphate on serine or threonine only, some catalyze the addition or hydrolysis of phosphate on tyrosine only, and

some catalyze the addition or hydrolysis of phosphate on serine, threonine, and tyrosine.

Membrane association is an important feature of signal transduction. Protein kinases propagate extracellular signals to the inside of the cell by attracting other signaling molecules to the membrane. Schlessinger and Ullrich, 1992, *Neuron* 9:383-391. For instance, many receptor protein kinases bind an extracellular ligand, dimerize, and cross phosphorylate one another. These phosphate moieties subsequently attract other proteins necessary for propagating the signal within the cell. The molecules that signal downstream of the receptor protein kinases are often non-receptor protein kinases which propagate and amplify the extracellular signal.

A class of non-receptor protein kinases are implicated in regulating RNA splicing. Fu, 1995 *RNA* 1:663-680; Staknis and Reed, 1994, *Mol. Cell. Biol.* 14:7670-7682. These protein kinases phosphorylate polypeptides rich in serine and arginine (SR proteins). SR proteins are characterized as containing at least one amino-terminal RNA recognition motif and a basic carboxy-terminal domain rich in serine and arginine residues, often arranged in tandem repeats. Zahler et al., 1992, *Genes Dev* 6:837-847. Experimental evidence supports the idea that the SR domain is involved in protein-protein interactions (Kohtz et al., 1994, *Nature* 368:119-124) as well as protein-RNA interactions (Harada et al., 1996,

Nature 380:175-179), and may contribute to a localization signal directing proteins to nuclear speckles. Hedley et al., 1995, *Proc. Natl. Acad. Sci. USA* 92:11524-11528.

A recent report demonstrated mCLK1, a CDC2 like kinase, interacts with ASF/SF2, SRp20 and hnRNP proteins in a yeast two hybrid system. Because hnRNP-K binds to the protooncogene p95^{vav}, mCLK1 could be implicated in transmitting signals that regulate the expression of the protooncogenes *myc* and *fos* in hematopoietic cells. Furthermore, it was demonstrated, that mCLK1 could phosphorylate ASF/SF2 *in vitro*, suggesting, that SR containing proteins are the natural substrates of mCLK1. Colwill et al., 1996, *EMBO J.* 15:265-275.

mCLK1 is a dual specificity protein kinase originally isolated in mouse expression libraries (Ben-David et al., 1991, *EMBO J.* 10:317-325; Howell et al., 1991, *Mol. Cell. Biol.* 11:568-572) and human (hCLK1, hCLK2, hCLK3), plant (AFC1, AFC2, AFC3) and fly (DOA) CLK protein kinases have since been identified. Johnson and Smith, 1991, *J. Biol. Chem.* 266:3402-3407; Hanes et al., 1994, *J. Mol. Biol.* 244:665-672; Bender and Fink, 1994, *Proc. Natl. Acad. Sci. USA* 91:12105-12109; Yun et al., 1994, *Genes. Dev.* 8:1160-1173. The amino terminal domain of these proteins is rich in serine and arginine, whereas the catalytic domain can be most similar to CDC2, a serine/threonine protein kinase. Ben-David et al., 1991, *EMBO J.* 10:317-325.

Both mCLK1 and the *Drosophila* homologue, DOA, regulate RNA splicing events. Each of these have two alternatively spliced products coding for either the full-length catalytically active protein or a truncated protein lacking the catalytic domain. Yun et al., 1994, *Genes. Dev.* 8:1160-1173; Duncan et al., 1995, *J. Biol. Chem.* 270:21524-21531. Identical splice forms were also found in human CLK protein kinases. Hanes et al., 1994, *J. Mol. Biol.* 244:665-672. The ratio of these splice products appears to be developmentally regulated in *Drosophila* (Yun et al., 1994, *Genes. Dev.* 8:1160-1173), and in a tissue and cell type specific manner in mammals. Hanes et al., 1994, *J. Mol. Biol.* 244:665-672; Duncan et al., 1995, *J. Biol. Chem.* 270:21524-21531. In addition, the expression of several other, larger transcripts, are observed to be differentially regulated and are shown to represent partially spliced products. Duncan et al., 1995, *J. Biol. Chem.* 270:21524-21531.

SUMMARY OF THE INVENTION

The present invention is based in part upon the isolation and characterization of nucleic acid molecules encoding CLK serine/threonine kinases designated mCLK2, mCLK3, and mCLK4. CLK serine/threonine kinases regulate RNA splicing in cells and some are highly expressed in cancer cells as well as testis. Various mCLK2, mCLK3, and mCLK4 related molecules and compounds can now be designed

as treatments of cancers or as contraceptives to reproduction in male organisms.

The present invention is based in part upon nucleic acid molecules encoding novel mCLK2, mCLK3, and mCLK4 polypeptides, nucleic acid molecules encoding portions of their amino acid sequences, nucleic acid vectors harboring such nucleic acid molecules, cells containing such nucleic acid vectors, purified polypeptides encoded by such nucleic acid molecules, and antibodies to such polypeptides, and methods of identifying compounds that bind mCLK2, mCLK3, and mCLK4 or abrogate their interactions with natural binding partners. Also disclosed are methods for diagnosing and treating specific abnormal conditions in an organism with mCLK2, mCLK3, and mCLK4 related molecules or compounds. The nucleic acid molecules, nucleic acid vectors, recombinant cells, polypeptides, and antibodies may be produced using well known and standard techniques used currently in the art.

Thus in a first aspect, the invention features isolated, enriched, or purified nucleic acid molecules encoding a novel mCLK2, mCLK3, or mCLK4 polypeptide.

The term "isolated", in reference to nucleic acid molecules, indicates that a naturally occurring sequence has been removed from its normal cellular environment. The isolated nucleic acid of the present invention is unique in the sense that it is not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been

removed from its normal cellular (i.e., chromosomal) environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide chain present, but that it is essentially free (about 90-95% pure at least) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

The term "enriched", in reference to nucleic acid molecules, means that the specific DNA or RNA sequence constitutes a significantly higher fraction (2 - 5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. A person skilled in the art could enrich a nucleic acid mixture by preferentially reducing the amount of other DNA or RNA present, or preferentially increasing the amount of the specific DNA or RNA, or both. However, nucleic acid molecule enrichment does not imply that there is no other DNA or RNA present, the term only indicates that the relative amount of the sequence of interest has been significantly increased. The term "significantly" qualifies "increased" to indicate that the level of increase is useful to the person performing the recombinant DNA technique, and generally means an increase relative to other nucleic acids of at least 2 fold, or more preferably at least 5 to 10 fold or more. The term also does not imply that there is no DNA or RNA

SD-80054.1

from other sources. Other DNA may, for example, comprise DNA from a yeast or bacterial genome, or a cloning vector. In addition, levels of mRNA may be naturally increased relative to other species of mRNA when working with viral infection or tumor growth techniques. This term distinguishes from naturally occurring events, such as viral infection, or tumor type growths, in which the level of one mRNA may be naturally increased relative to other species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation). Instead, it represents an indication that the sequence is relatively more pure than in the natural environment (compared to the natural level this level should be at least 2-5 fold greater, e.g., in terms of mg/mL). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the

SD-80054.1

creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process, which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones, yields an approximately 10^6 -fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

The term "nucleic acid molecule" describes a polymer of deoxyribonucleotides (DNA) or ribonucleotides (RNA). The nucleic acid molecule may be isolated from a natural source by cDNA cloning or subtractive hybridization or synthesized manually. The nucleic acid molecule may be synthesized manually by the triester synthetic method or by using an automated DNA synthesizer.

"cDNA cloning" techniques refer to hybridizing a small nucleic acid molecule, a probe, to genomic cDNA that is bound to a membrane. The probe hybridizes (binds) to complementary sequences of cDNA. The term "complementary" describes two nucleotides that can form multiple favorable interactions with one another. For example, adenine is complementary to thymidine as they can form two hydrogen bonds. Similarly, guanine and cytosine are complementary since they can form three hydrogen bonds. cDNAs are molecules that are reverse transcribed from fragments of

message RNA from a genomic source. These fragments form a cDNA library of nucleic acid molecules. cDNA libraries are constructed from natural sources such as mammalian blood, semen, or tissue.

The term "subtractive hybridization" refers to a method similar to cDNA-cloning except that cDNA prepared from mRNA in unstimulated cells is added to mRNA in stimulated or different types of cells. cDNA/mRNA can then be precipitated to enrich the mRNA specific to the stimulation signal or different cell.

The term "hybridize" refers to a method of interacting a nucleic acid probe with a DNA or RNA molecule in solution or on a solid support, such as cellulose or nitrocellulose. If a nucleic acid probe binds to the DNA or RNA molecule with high affinity, it is said to "hybridize" to the DNA or RNA molecule. As mentioned above, the strength of the interaction between the probe and its target can be assessed by varying the stringency of the hybridization conditions. Various low or high stringency hybridization conditions may be used depending upon the specificity and selectivity desired. Stringency is controlled by varying salt or denaturant concentrations. Under stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having one or two mismatches out of 20 contiguous nucleotides.

By "novel" is meant new and in the particular context of the present invention refers to CLK sequences that have not been previously described. In preferred embodiments the novel sequence may be the full-length serine CLK2 or CLK3 sequence, the full-length mammalian CLK4 sequence, or shorter fragments (preferably functional fragments) if any of the above as long as they were not already previously described.

The terms "mCLK2", "mCLK3", and "mCLK4" refer to polypeptides that have amino acid sequences substantially similar to those set forth in Figure 1, Figure 2, Figure 4, or Figure 6. A sequence that is substantially similar will preferably have at least 95% identity, more preferably at least 96-97% identity, and most preferably 98-100% identity to the sequence set forth in Figure 1, Figure 2, Figure 4, or Figure 6. CLK protein kinase polypeptides preferably have protein kinase activity and fragments of the full length CLK protein kinase sequences having such activity may be identified using techniques well known in the art, such as sequence comparisons and assays such as those described in the examples herein. Other aspects of mCLK2, mCLK3, and mCLK4 nucleic acid sequences, amino acid sequences, functions and properties are further depicted in Nayler et al., 1997, Biochem J. 326: 693-700, hereby incorporated by reference herein in its entirety including all figures, tables, and drawings.

By "identity" is meant a property of sequences that measures their similarity or relationship. Generally

speaking, identity is measured by dividing the number of identical residues by the total number of residues and gaps and multiplying the product by 100. "Gaps" are spaces in an alignment that are the result of additions or deletions of amino acids. Thus, two copies of exactly the same sequence have 100% identity, but sequences that are less highly conserved, and have deletions, additions, or replacements, may have a lower degree of identity. Those skilled in the art will recognize that several computer programs are available for determining sequence identity.

A preferred embodiment of the invention concerns nucleic acid molecules relating to mCLK2, mCLK3, and mCLK4 that are enriched, isolated, or purified from a mammalian source. These nucleic acid molecules can be isolated from, among others, blood, semen, or tissue. Although mCLK2, mCLK3, and mCLK4 nucleic acid molecules are isolated from mouse cells, current recombinant DNA techniques can readily elucidate related nucleic acid molecules in other mammalian tissue. Mammals include, but are not limited to, mice, rats, rabbits, cows, horses, monkeys, apes, and preferably humans.

Another preferred embodiment of the invention concerns isolated nucleic acid molecules that encode at least seventeen amino acids of a mCLK2, mCLK3, or mCLK4 polypeptide. Preferably, at least 17, 20, 25, 30, 35, 40, 50, 100, 200, 300, 400, 450, 475, or 485 contiguous amino acids are encoded. This preferred embodiment of the

invention is achieved by applying routine recombinant DNA techniques known to those skilled in the art.

Another aspect of the invention features a nucleic acid probe that can detect nucleic acid molecules encoding a mCLK2, mCLK3, or mCLK4 polypeptide in a sample.

The term "nucleic acid probe" refers to a nucleic acid molecule that is complementary to and can bind a nucleic acid sequence encoding the amino acid sequence substantially similar to that set forth in Figure 1, Figure 2, Figure 4, or Figure 6.

The nucleic acid probe or its complement encodes any one of the amino acid molecules set forth in the invention. Thus the nucleic acid probe can encode at least 17, 20, 25, 30, 35, 40, 50, 100, 200, 300, 400, 450, 475, or 485 contiguous amino acids of the full-length sequence set forth in Figure 1, Figure 2, Figure 4, or Figure 6.

The nucleic acid probe can be labeled with a reporter molecule or molecules. The term "reporter molecule" refers to a molecule that is conjugated to the nucleic acid probe or is contained within the nucleic acid probe. The reporter molecule allows the detection of the probe by methods used in the art. Reporter molecules are chosen from, but not limited to, the group consisting of an enzyme, such as a peroxidase, a radioactive element, or an avidin or biotin molecule.

A nucleic acid probe, whether labeled or unlabeled, should hybridize to a complement in a sample. Various low

or high stringency hybridization conditions may be used depending upon the specificity and selectivity desired. Stringency is controlled by varying salt or denaturant concentrations. Under stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acid molecules having one or two mismatches out of 20 contiguous nucleotides, more preferably prevent hybridization of nucleic acid molecules having one mismatch out of 35 contiguous nucleotides, and most preferably prevent hybridization of nucleic acid molecules having one mismatch out of 50 contiguous nucleotides.

The nucleic acid probe or complement can also refer to a nucleic acid molecule encoding a conserved or unique region of amino acids. These nucleic acid molecules are useful as hybridization probes to identify and clone additional polypeptides relating to CLK serine/threonine kinases. The term "conserved nucleic acid regions" refers to regions present in two or more nucleic acid molecules encoding a CLK protein kinase polypeptide, to which a particular nucleic acid sequence can hybridize under lower stringency conditions. Examples of lower stringency conditions suitable for screening nucleic acid molecules are provided in Abe, et al. J. Biol. Chem., 19:13361 (1992) (hereby incorporated by reference herein in its entirety, including any drawings). Preferably, conserved regions differ by no more than 5 out of 20 nucleotides,

more preferably conserved regions differ by no more than 10 out of 20 nucleotides, and most preferably conserved regions differ by no more than 15 out of 20 nucleotides. Protein kinases share conserved regions in the catalytic domain.

The term "unique nucleic acid region" concerns a sequence present in a full length nucleic acid molecule encoding a CLK protein kinase polypeptide that is not present in a sequence encoding any other naturally occurring polypeptide. Such regions preferably comprise 30 or 45 contiguous nucleotides, more preferably 100 contiguous nucleotides, and most preferably 200 contiguous nucleotides present in the full length nucleic acid sequence encoding a CLK protein kinase polypeptide. In particular, a unique nucleic acid region is preferably of mammalian origin.

Methods for using the probes include detecting the presence or amount of CLK protein kinase RNA in a sample by contacting the sample with a nucleic acid probe under conditions such that hybridization occurs and detecting the presence or amount of the probe bound to CLK RNA. The nucleic acid duplex formed between the probe and a nucleic acid sequence encoding a CLK protein kinase polypeptide may be used in the identification of the sequence of the nucleic acid detected (for example see, Nelson et al., in Nonisotopic DNA Probe Techniques, p. 275 Academic Press, San Diego (Kricka, ed., 1992) hereby incorporated by reference herein in its entirety, including any drawings).

Kits for performing such methods may be constructed to include a container holding a nucleic acid probe.

In yet another aspect, the invention relates to a nucleic acid vector comprising a promoter element and a nucleic acid molecule described in the first aspect of the invention.

The term "nucleic acid vector" relates to a single or double stranded circular nucleic acid molecule that can be transfected or transformed into cells and replicate independently or within a cell genome. A vector can be cut and thereby linearized upon treatment with restriction enzymes. An assortment of vectors, restriction enzymes, and the knowledge of the nucleotide sequences that the restriction enzymes operate upon are readily available to those skilled in the art. A nucleic acid molecule encoding a CLK protein kinase can be inserted into a vector by cutting the vector with restriction enzymes and ligating the two pieces together.

The term "promoter element" describes a nucleotide sequence that is incorporated into a vector that, once inside an appropriate cell, may facilitate transcription factor and/or polymerase binding and subsequent transcription of portions of the vector DNA into mRNA. The promoter element precedes the 5' end of the nucleic acid molecule of the first aspect of the invention such that the latter is transcribed into mRNA. Recombinant cell machinery then translates mRNA into a polypeptide.

Many techniques are available to those skilled in the art to facilitate transformation or transfection of the nucleic acid vector into a prokaryotic or eukaryotic organism. The terms "transformation" and "transfection" refer to methods of inserting a nucleic acid vector into a cellular organism. These methods involve a variety of techniques, such as treating the cells with high concentrations of salt, an electric field, or detergent, to render the cell outer membrane or wall permeable to nucleic acid molecules of interest.

A nucleic acid vector can be useful for identifying natural binding partners of CLK serine/threonine kinases.

The term "natural binding partners" refers to polypeptides that bind to CLK serine/threonine kinases and play a role in propagating a signal in a signal transduction process. The term "natural binding partner" also refers to a polypeptide that binds to CLK serine/threonine kinases within a cellular environment with high affinity. High affinity represents an equilibrium binding constant on the order of 10^{-1} M. However, a natural binding partner can also transiently interact with a CLK protein kinase and chemically modify it. CLK protein kinase natural binding partners are chosen from a group consisting of, but not limited to, src homology 2 (SH2) or 3 (SH3) domains, other phosphoryl tyrosine binding domains, and receptor and non-receptor protein kinases or protein, phosphatases.

Methods are readily available in the art for identifying binding partners of polypeptides of interest. These methods include screening cDNA libraries included in one nucleic acid vector with a nucleic acid molecule encoding the desired polypeptide in another nucleic acid vector. Wojtek et al., 1993, Cell 74:205214. These techniques often utilize yeast recombinant cells. These techniques also utilize two halves of a transcription factor, one half that is fused to a polypeptide encoded by the cDNA library and the other that is fused to the polypeptide of interest. Interactions between a polypeptide encoded by the cDNA library and the polypeptide of interest are detected when their interaction concomitantly brings together the two halves into an active transcription factor which in turn activates a gene that reports the interaction. Any of the nucleic molecules encoding mCLK2, mCLK3, or mCLK4 can be readily incorporated into an nucleic acid vector used in such a screening procedure by utilizing standard recombinant DNA techniques in the art.

Another aspect of the invention relates to a recombinant cell or tissue comprising a nucleic acid molecule encoding a mCLK2, mCLK3, or mCLK4 polypeptide.

The term "recombinant" refers to an organism that has a new combination of genes or nucleic acid molecules. A new combination of genes or nucleic acid molecules can be introduced to an organism using a wide array of nucleic

acid manipulation techniques available to those skilled in the art.

The recombinant cell can be a eukaryotic or prokaryotic organism. The term "eukaryote" refers to an organism comprised of cells containing a nucleus. Eukaryotes are differentiated from "prokaryotes" which do not house their genomic DNA inside a nucleus. Prokaryotes include unicellular organisms such as bacteria while eukaryotes are represented by yeast, invertebrates, and vertebrates.

The recombinant cell can also harbor a nucleic acid vector that is extragenomic. The term "extragenomic" refers to a nucleic acid vector which does not integrate into a cell genome. Many nucleic acid vectors are designed with their own origins of replication which allow them to utilize the recombinant cell replication machinery to copy and propagate the nucleic acid vector nucleic acid sequence. These nucleic acid vectors are small enough that they are not likely to harbor nucleic acid sequences homologous to genomic sequences of the recombinant cell. Thus these nucleic acid vectors replicate independently of the genome and do not recombine with or integrate into the genome.

A recombinant cell can also harbor a portion of a nucleic acid vector in an intragenomic fashion. The term "intragenomic" defines a nucleic acid vector that integrates within a cell genome. Multiple nucleic acid vectors available to those skilled in the art contain

nucleic acid sequences that are homologous to nucleic acid sequences in a particular organism's genomic DNA. These homologous sequences will result in recombination events that incorporate portions of the nucleic acid vector into the genomic DNA. Those skilled in the art can control which nucleic acid sequences of the nucleic acid vector integrate into the cell genome by flanking the portion to be integrated into the genome with homologous sequences in the nucleic acid vector.

In yet another aspect, the invention features an isolated, enriched, or purified polypeptide encoded by a mCLK2, mCLK3, or mCLK4 nucleic acid molecule of the invention.

The term "isolated", in reference to a polypeptide, describes a polymer of amino acids conjugated to each other that are separated from a natural source. The polypeptide can also be synthesized manually. Isolated peptides can be at least 17, 20, 25, 30, 35, 40, 50, 100, 200, or 300 contiguous amino acids of one of the full-length sequences set forth in Figure 1, Figure 2, Figure 4, or Figure 6. In certain aspects longer polypeptides are preferred, such as those with 400, 450, 475, or 485 of the contiguous amino acids of mCLK2, mCLK3, or mCLK4 set forth in Figure 1, Figure 2, Figure 4, or Figure 6.

The isolated polypeptides of the present invention are unique in the sense that they are not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been

removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only amino acid chain present, but that it is essentially free (about 90 - 95% pure at least) of non-amino acid material naturally associated with it.

The term "enriched", in reference to a polypeptide, defines a specific amino acid sequence constituting a significantly higher fraction (2 - 5 fold) of the total of amino acids present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was separated. A person skilled in the art can preferentially reduce the amount of other amino acid sequences present, or preferentially increase the amount of specific amino acid sequences of interest, or both. However, the term "enriched" does not imply that there are no other amino acid sequences present. Enriched simply means the relative amount of the sequence of interest has been significantly increased. The term "significant" indicates that the level of increase is useful to the person making such an increase. The term also means an increase relative to other amino acids of at least 2 fold, or more preferably at least 5 to 10 fold, or even more. The term also does not imply that there are no amino acid sequences from other sources. Other source amino acid sequences may, for example, comprise amino acid sequences from a recombinant organism.

"Enriched" is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired amino acid sequence.

The term "purified", in reference to a polypeptide, does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the amino acid sequence is relatively more pure than in a cellular environment. The concentration of the preferred amino acid sequence should be at least 2-5 fold greater (in terms of mg/ml) than its concentration in a cellular environment. Purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is preferred. The substance is preferably free of contamination, as indicated by purity levels of 90%, 95%, or 99%.

A preferred embodiment of the invention relates to a mCLK2, mCLK3, or mCLK4 polypeptide that is a unique fragment. This unique fragment can contain at least 17, 20, 25, 30, 35, 40, 50, 100, 200, 300, 400, 450, 475, or 485 contiguous amino acids of one of the full-length sequences. In addition, preferred lengths and portions of mCLK2, mCLK3, or mCLK4 amino acid sequences are encoded by the nucleic acid molecules defined in the first aspect of the invention.

The term "unique fragment" refers to the minimum stretch of amino acids in one mCLK molecule that is different in sequence than any other portion of another

protein kinase. Since the largest identical stretch of amino acids found in Figure 1, Figure 2, Figure 4, or Figure 6 is seventeen amino acids, the minimum unique fragment for a mCLK protein kinase is seventeen amino acids.

The polypeptide can be isolated, enriched, or purified from a prokaryotic or eukaryotic recombinant cell. A eukaryotic cell includes mammals and preferably humans. Multiple standard techniques are available to those skilled in the art to facilitate isolation, enrichment, or purification of a polypeptide from recombinant cells. These methods typically include lysing the recombinant cells and separating the polypeptide of interest from the rest of the cell polypeptides, nucleic acids, and fatty acid-based material using standard chromatography techniques known in the art.

Another aspect of the invention features an antibody, that is monoclonal or polyclonal, or an antibody fragment having specific binding affinity to a mCLK2, mCLK3, or mCLK4 polypeptide.

Antibodies or antibody fragments are polypeptides with regions that can bind to other polypeptides with high affinity. The term "specific binding affinity" describes an antibody that binds to a mCLK2, mCLK3, or mCLK4 polypeptide with greater affinity than it binds to other polypeptides under specified conditions.

The term "polyclonal" refers to a mixture of antibodies with specific binding affinity to a mCLK2,

mCLK3, or mCLK4 polypeptide, while the term "monoclonal" refers to one antibody with specific binding affinity to a mCLK2, mCLK3, or mCLK4 polypeptide. A monoclonal antibody binds to one specific region on a mCLK2, mCLK3, or mCLK4 polypeptide and a polyclonal mixture of antibodies can bind multiple regions of a mCLK2, mCLK3, or mCLK4 polypeptide. One skilled in the art would note that a monoclonal and especially a polyclonal antibody that has specific binding affinity to a mCLK2, mCLK3, or mCLK4 polypeptide will most likely also have specific binding affinity to another CLK protein kinase polypeptide of mammalian origin.

The term "antibody fragment" refers to a portion of an antibody, often the hypervariable region and portions of the surrounding heavy and light chains, that displays specific binding affinity for a particular molecule. A hypervariable region is a portion of an antibody that physically binds to the ligand to which it binds specifically.

Antibodies or antibody fragments having specific binding affinity to a mCLK2, mCLK3, or mCLK4 polypeptide may be used in methods for detecting the presence and/or amount of a CLK protein kinase polypeptide in a sample by probing the sample with the antibody under conditions suitable for CLK protein kinase-antibody immunocomplex formation and detecting the presence and/or amount of the antibody conjugated to a CLK protein kinase polypeptide. Diagnostic kits for performing such methods may be

constructed to include antibodies or antibody fragments specific for a CLK protein kinase as well as a conjugate of a binding partner of the antibodies or the antibodies themselves.

Another aspect of the invention features a hybridoma which produces an antibody having specific binding affinity to a mCLK2, mCLK3, or mCLK4 polypeptide. A "hybridoma" is an immortalized cell line which is capable of secreting an antibody, for example an antibody with specific binding affinity to a mCLK2, mCLK3, or mCLK4 polypeptide.

Another aspect of the invention relates to an isolated, enriched, or purified nucleic acid molecule comprising a nucleotide sequence that: (a) encodes a full length mCLK2, mCLK3, or mCLK4 amino acid sequence as set forth in Figure 1, Figure 2, Figure 4, or Figure 6; (b) encodes the complement of the nucleotide sequence encoding the amino acid sequences of Figure 1, Figure 2, Figure 4, or Figure 6; (c) hybridizes under highly stringent conditions to the nucleic acid molecule of (a) and encodes a naturally occurring mCLK2, mCLK3, or mCLK4 protein; (d) a mCLK2, mCLK3, or mCLK4 protein having the full length amino acid sequence as set forth in Figure 1, Figure 2, Figure 4, or Figure 6 except that it lacks one or more of the following segments of amino acid residues 1-182, 183-470, or 471-499 of mCLK2, 1-176, 177-473, or 474-496 of mCLK3, or 1183, 184-486, or 486-489 of mCLK4; (e) the complement of the nucleotide sequence of (d); (f)

a polypeptide having the amino acid sequence set forth in Figure 1, Figure 2, Figure 4, or Figure 6 from amino acid residues 1-182, 183-470, or 471-499 of mCLK2, 1176, 177-473, or 474-496 of mCLK3, or 1-183, 184-486, or 486-489 of mCLK4; (g) the complement of the nucleotide sequence of (f); (h) encodes a polypeptide having the full length amino acid sequence set forth in Figure 1, Figure 2, Figure 4, or Figure 6 except that it lacks one or more of the domains selected from the group consisting of a N-terminal domain, a catalytic domain, and a C-terminal-region; or (i) the complement of the nucleotide sequence of (h).

The term "N-terminal domain" refers to a portion of the full length mCLK2, mCLK3, or mCLK4 amino acid sequences spanning from the amino terminus to the start of the catalytic domain.

The term "catalytic domain" refers to a portion of the full length mCLK2, mCLK3, or mCLK4 amino acid molecules that does not contain the N-terminal domain or C-terminal region and has catalytic activity.

The term "C-terminal region" refers to a portion of the full length mCLK2, mCLK3, or mCLK4 amino acid molecules that begins at the end of the catalytic domain and ends at the carboxy terminal amino acid, which is the last amino acid encoded before the stop codon in the nucleic acid sequence.

Domains are regions of polypeptides which contain particular functions. For instance, N-terminal or C-

terminal domains of signal transduction proteins can serve functions including, but not limited to, binding molecules that localize the signal transduction molecule to different regions of the cell or binding other signaling molecules directly responsible for propagating a particular cellular signal. Some domains can be expressed separately from the rest of the protein and function by themselves, while others must remain part of the intact protein to retain function. The latter are termed functional regions of proteins and also relate to domains.

Functional regions of mCLK2, mCLK3, or mCLK4 may be identified by aligning their amino acid sequences with amino acid sequences of other polypeptides with known functional regions. If regions of mCLK2, mCLK3, or mCLK4 share high amino acid identity with the amino acid sequences of known functional regions, then mCLK2, mCLK3, or mCLK4 can be determined to contain these functional regions by those skilled in the art. The functional regions can be determined, for example, by using computer programs and sequence information available to those skilled in the art.

Other functional regions of signal transduction molecules that may exist within mCLK2, mCLK3, or mCLK4 include, but are not limited to, proline-rich regions or phosphoryl tyrosine regions. These regions can interact with natural binding partners such as SH2 or SH3 domains of other signal transduction molecules.

Another aspect of the invention relates to nucleic acid vectors comprising any of the nucleic acid molecules described herein.

In another aspect, the invention includes recombinant cells or tissues comprising any of the nucleic acid molecules described herein.

In yet another aspect, the invention relates to a method of identifying compounds capable of inhibiting or activating CLK protein kinase phosphorylation activity. This method comprises the following steps: (a) adding a compound to a mixture comprising a CLK protein kinase polypeptide and a substrate for a CLK protein kinase; and (b) detecting a change in phosphorylation of said substrate.

The term "compound" includes small organic molecules including, but not limited to, oxindolinones, quinazolines, tyrphostins, quinoxalines, and extracts from natural sources.

The term "CLK protein kinase polypeptide" refers to any CLK protein kinase isolated from an insect or a mammal. The polypeptide can be the full length amino acid sequence (the contiguous amino acids encoded by the nucleic acids spanning from the start codon to the stop codon of a naturally occurring CLK protein kinase nucleic acid molecule) or portions of a naturally occurring full length CLK protein kinase. Preferably, at least 17, 20, 25, 30, 35, 40, 50, 100, 200, 300, 400, 450, 475, or 485

contiguous amino acids are encoded for the CLK protein kinase polypeptide.

The term "a change in phosphorylation", in the context of the invention, defines a method of observing a change in phosphorylation of the substrate in response to adding a compound to cells. The phosphorylation can be detected, for example, by measuring the amount of a substrate that is converted to a product with respect to time. Addition of a compound to cells expressing a CLK protein kinase polypeptide may either enhance (activate) or lower (inhibit) the phosphorylation. If a compound lowers phosphorylation, the compound is assumed to bind to a CLK protein kinase polypeptide and block the ability of CLK protein kinase to bind and/or turn over a substrate. If a compound enhances phosphorylation, the compound is assumed to bind to a CLK protein kinase polypeptide and facilitate the ability of CLK protein kinase to bind and/or turn over a substrate.

The method can utilize any of the molecules disclosed in the invention. These molecules include nucleic acid molecules encoding mCLK2, mCLK3, or mCLK4 polypeptides, nucleic acid vectors, recombinant cells, polypeptides, or antibodies of the invention.

Another aspect of the invention is a method of identifying compounds useful for the diagnosis or treatment of an abnormal condition in an organism. The abnormal condition can be associated with an aberration in a signal transduction pathway characterized by an

interaction between a CLK protein kinase polypeptide and a natural binding partner. The method comprises the following steps: (a) adding a compound to cells; and (b) detecting whether the compound promotes or disrupts said interaction between a CLK protein kinase polypeptide and a natural binding partner.

The term "abnormal condition" refers to a function in an organism's cells or tissue that deviate from a normal function in the cells or tissue of that organism. In the context of this aspect of the invention, abnormal conditions can be associated with cell proliferation or with RNA splicing.

Aberrant cell proliferative conditions include cancers such as fibrotic and mesangial disorders, abnormal angiogenesis and vasculogenesis, wound healing, psoriasis, diabetes mellitus, and inflammation.

RNA splicing is a necessary function of a cell that occurs in a cell nucleus. This process is the last step in the synthesis of message RNA from DNA. One molecule of RNA transcribed from DNA is tied into a lariat, incised in at least two places at the intersection of the strands, the lariat is excised, and the non-excised portion is ligated together. The modified RNA is then fit to be message RNA and is ejected from the cell nucleus to be translated into a polypeptide. Thus any aberrations that exist in an organisms ability to splice the RNA of a particular gene could result in the deficiency of a cellular agent and give rise to an abnormal condition.

Thus, regulating RNA splicing could be useful in treating cancer. For example, it is known that proteins such as Raf or src become oncogenic when made in a truncated form, such as could happen when RNA is incorrectly spliced. For this reason, the proteins of the invention might be useful for finding compounds to treat cancer. In addition, molecules involved in RNA processing have been linked to spermatogenesis. Thus, modifying RNA processing could lead to more sperm (to treat infertility) or less sperm. These methods would preferably involve CLK3 due to its high expression in the testis.

The abnormal condition can be diagnosed when the organism's cells exist within the organism or outside of the organism. Cells existing outside the organism can be maintained or grown in cell culture dishes. For cells harbored within the organism, many techniques exist in the art to administer compounds, including (but not limited to) oral, parenteral, dermal, and injection applications. For cells outside of the patient, multiple techniques exist in the art to administer the compounds, including (but not limited to) cell microinjection techniques, transformation techniques, and carrier techniques.

The term "signal transduction pathway" refers to the molecules that propagate an extracellular signal through the cell membrane to become an intracellular signal. This signal can then stimulate a cellular response. The polypeptide molecules involved in signal transduction

processes are typically receptor and non-receptor protein kinases, receptor and non-receptor protein phosphatases, and transcription factors.

The term "aberration", in conjunction with a signal transduction process, refers to a CLK protein kinase polypeptide that is over- or under-expressed in an organism, mutated such that its catalytic activity is lower or higher than wild-type CLK protein kinase, mutated such that it can no longer interact with a binding partner, is no longer modified by another protein kinase or protein phosphatase, or no longer interacts with a binding partner.

The term "interaction" defines the complex formed between a CLK protein kinase polypeptide and a natural binding partner. Compounds can bind to either the CLK protein kinase polypeptide or the natural binding partner and disrupt the interaction between the two molecules. The method can also be performed by administering a group of cells containing an aberration in a signal transduction process to an organism and monitoring the effect of administering a compound on organism function. The art contains multiple methods of introducing a group of cells to an organism as well as methods of administering a compound to an organism. The organism is preferably an animal such as a frog, mouse, rat, rabbit, monkey, or ape, and also a human.

Methods of determining a compound's effect of detecting an interaction between CLK serine/threonine

kinases and natural binding partners exist in the art. These methods include, but are not limited to, determining the effect of the compound upon the catalytic activity of a CLK protein kinase polypeptide, the phosphorylation state of the CLK protein kinase polypeptides or natural binding partners, the ability of a CLK protein kinase to bind a natural binding partner, or a difference in a cell morphology.

Differences in cell morphology include growth rates, differentiation rates, cell hypertrophy, survival, or prevention of cell death. These phenomena are simply measured by methods in the art. These methods can involve observing the number of cells or the appearance of cells under a microscope with respect to time (days).

Another aspect of the invention relates to a method of diagnosing an abnormal condition associated with cell proliferation or RNA splicing in an organism. The abnormal condition can be associated with an aberration in a signal transduction pathway characterized by an interaction between a CLK protein kinase polypeptide and a natural binding partner. The method comprises the step of detecting the abnormal interaction.

The abnormal interaction can be assessed by the methods described above in reference to the identification of compounds useful for diagnosing an abnormal condition in an organism.

In a final aspect, the invention features a method of administering a compound to a male organism that acts a

contraceptive to reproduction. The compound can inhibit the catalytic activity of a CLK protein kinase or inhibit the binding of a natural binding partner to a CLK protein kinase.

Preferred embodiments of the methods of the invention relate to CLK serine/threonine kinases that are isolated from mammals, preferably humans, and to organisms that are mammals, preferably humans.

The summary of the invention described above is not limiting and other features and advantages of the invention will be apparent from the following detailed description of the invention, and from the claims.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 compares amino acid sequences encoded by mCLK1, mCLK2, mCLK3, and mCLK4 nucleic acid molecules cloned from mouse cells. Each amino acid sequence is encoded between a start codon and a stop codon from its respective nucleic acid molecule. Dots indicate identical amino acids and hyphens are introduced for optimal alignment. The predicted nuclear localization signals are underlined. Invariant amino acids signifying CDC2 like kinases are printed in bold. The catalytic domain is indicated by arrows. The LAMMER signature is indicated by asterisks.

Figure 2 depicts an amino acid sequence of mCLK2.

Figure 3 depicts a nucleic acid sequence of mCLK2.

Figure 4 depicts an amino acid sequence of mCLK3.

Figure 5 depicts a nucleic acid sequence of mCLK3.

Figure 6 depicts an amino acid sequence of mCLK4.

Figure 7 depicts a nucleic acid sequence of mCLK4.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based in part upon the isolation and characterization of nucleic acid molecules encoding novel protein kinases designated mCLK2, mCLK3, and mCLK4. The invention also relates to nucleic acid molecules encoding portions of these protein kinase polypeptides, nucleic acid molecules encoding at least one mCLK2, mCLK3, or mCLK4 functional region, nucleic acid vectors harboring such nucleic acid molecules, cells containing such nucleic acid vectors, purified polypeptides encoded by such nucleic acid molecules, antibodies to such polypeptides, and methods of identifying compounds that bind CLK serine/threonine kinases or abrogate their interactions with natural binding partners. Also disclosed are methods for diagnosing abnormal conditions in an organism with CLK protein kinase related molecules or compounds. The invention also concerns using a CLK protein kinase related molecule or compound as a contraceptive to reproduction in a male organism.

The present invention discloses the discovery of the protein kinases, mCLK2, mCLK3, and mCLK4. The predicted molecular weights of the encoded proteins are 59.9kDa (mCLK2), 58.5kDa (mCLK3), and 57.2kDa (mCLK4).

As illustrated in Figure 1, mCLK1, mCLK2, mCLK3, and mCLK4 share the essential features identifying them as LAMMER kinases. Yun et al., 1994, *Genes. Dev.* 8:1160-1173. They contain a nuclear localization signal (Dingwall and Laskey, 1991, *Trends Biochem. Sci.* 16:478-481), as well as an unusually basic amino terminus composed of many serine and arginine residues. These serine and arginine amino acids likely embody a signal sequence localizing the protein to nuclear speckles. Hedley et al., 1995, *Proc. Natl. Acad. Sci. USA* 92:11524-11528; Colwill et al., 1996, *EMBO J.* 15:265-275. The amino terminus is the most divergent portion of the proteins, suggesting that this area could contain information specific to each protein. The catalytic domain is homologous among all family members, with only few amino acid changes. Furthermore, all amino acids known to define the subfamily of CDC2 like kinases are present in all four proteins. Ben-David et al., 1991, *EMBO J.* 10:317-325.

mCLK1 has been shown to interact with ASF/SF2, SRp20 and hnRNP proteins in a yeast two hybrid system. Because hnRNP-K binds to the protooncogene p95^{vav}, mCLK1 could be implicated in transmitting signals that regulate the expression of the protooncogenes *myc* and *fos* in hematopoietic cells. Thus the role of CLK serine/threonine kinases may not be limited to simply maintaining RNA splicing and translocation events in the cell; CLK serine/threonine kinases may also be linked to

regulating the flow of extracellular signals within hematopoietic cells. In addition, CLK serine/threonine kinases may be targets for compounds that could ameliorate cancers associated with uncontrolled regulation of the protooncogenes *p95^{vav}*, *myc*, and *fos*. Because over-expression of CLK serine/threonine kinases themselves have been implicated in certain types of cancer cell lines, compounds that inhibit their catalytic activity or disrupt their interactions with natural binding partners may act as anti-cancer therapeutics.

Even though CLK serine/threonine kinases other than mCLK2, mCLK3, and mCLK4 have been described previously, the methods of the invention relate to CLK serine/threonine kinases in general as the methods described herein are not disclosed elsewhere. Thus the methods of the invention include antibodies and other compounds with specific binding affinity to mCLK2, mCLK3, and mCLK4 as well as antibodies and other compounds that interact with other CLK protein kinase polypeptides.

Various other features and aspects of the invention are: nucleic acid molecules encoding a mCLK2, mCLK3, or mCLK4 polypeptide; nucleic acid probes for the detection of CLK serine/threonine kinases; a probe-based method and kit for detecting CLK protein kinase messages in other organisms; DNA constructs comprising a mCLK2, mCLK3, or mCLK4 nucleic acid molecule and cells containing these constructs; purified mCLK2, mCLK3, or mCLK4 polypeptides; mCLK2, mCLK3, or mCLK4 antibodies and hybridomas;

antibody-based methods and kits for detecting CLK serine/threonine kinases; identification of agents; isolation of compounds which interact with a CLK protein kinase polypeptide; compositions of compounds that interact with CLK serine/threonine kinases; pharmaceutical formulations and modes of administration; derivatives of complexes; antibodies to complexes; disruption of CLK protein kinase protein complexes; purification and production of complexes; transgenic animals containing mCLK2, mCLK3, or mCLK4 nucleic acid constructs; antisense and ribozyme approaches, gene therapy; and evaluation of disorders. one skilled in the art would note that a derivative of a complex can manifest itself as a derivative of any one of the molecules in that complex, including derivatives of a nucleic acid molecule, a polypeptide, or a compound bound to a polypeptide. All of these aspects and features are explained in detail with respect to the protein PYK-2 in PCT publication WO 96/18738, which is incorporated herein by reference in its entirety, including any drawings. Those skilled in the art will readily appreciate that such descriptions can be easily adapted to mCLK2, mCLK3, mCLK4, or other CLK serine/threonine kinases as well, and is equally applicable to the present invention.

Other features and aspects of the invention are depicted in PCT Application WO 97/48723 (PCT/IB97/00946), published on December 24, 1997, filed on June 17, 1997, Axel Ullrich et al., entitled "Novel PTP20, PCP-2, BDP-1, SD-80054.1

CLK, and SIRP Proteins and Related Products and Methods," hereby incorporated by reference herein in its entirety, including all figures, tables, and drawings.

EXAMPLES

The examples below are non-limiting and are merely representative of various aspects and features of the present invention. The examples below demonstrate the isolation, and characterization of the novel protein kinases, mCLK2, mCLK3, and mCLK4.

EXAMPLE 1: PCR AMPLIFICATION AND CLONING

The catalytic domain of protein kinases contains highly conserved regions, which have been successfully used to PCR amplify and clone novel family members from a variety of species and tissues. The signature sequence HRDLAAR in the catalytic subdomain VI and D(V/M)WS(Y/F)G in subdomain IX were used to create degenerate oligonucleotides. Ciossek et al., 1995, *Oncogene* 11:2085-2095. These primers were then used to search for unknown protein kinases involved in muscle cell differentiation using reverse transcriptase PCR of total RNA, isolated from various *in vitro* differentiated stages of the mouse myoblast cell line C2C12. Lechner et al., 1996, *Proc. Natl. Acad. Sci. USA* 93:4355-4359.

From the approximately 300 fragments which were sequenced one was novel. It derived from a member of the LAMMER family of dual specificity kinases (Yun et al.,

1994, *Genes. Dev.* 8:1160-1173), also known as CLK kinases (Ben-David et al., 1991, *EMBO J.* 10:317-325) or STY (Howell et al., 1991, *Mol. Cell. Biol.* 11:568-572) and shared a high homology to a part of the human cDNA hCLK2. To obtain the full length clone and to search for other closely related sequences, a mouse 11.5 p.c. embryonic library was screened at low stringency using the original 250bp PCR fragment as a probe. Three highly related full-length cDNA sequences defining different members of the CLK family using this technique.

The same libraries were rescreened with a mixture of mCLK1, 2, 3, and 4 fragments at low stringency to isolate additional novel members of this family.

Reverse transcriptase PCR reactions were performed on brain, kidney and liver poly (A)⁺ RNA with degenerate primers coding for the DLKPEN (SEQ ID NO. 1) and AMMERI (SEQ ID NO. 2) motifs. These efforts did not identify additional genes.

Reverse transcriptase PCR reactions were performed with 2µg of total RNA prepared from confluent or differentiated (day 7) mouse C2C12 myoblasts (Lechner et al., 1996, *Proc. Natl. Acad. Sci. USA* 93:4355-4359) using degenerate oligonucleotide primers. Ciossek et al., 1995, *Oncogene* 11:2085-2095. Briefly, 2µg of RNA were reverse transcribed in the presence of 1µM degenerate antisense primer, 250µM of each nucleotide and 75 units of Stratascript reverse transcriptase (Stratagene) in a total volume of 20µl for 30 min at 42°C. 2µl of the above

reaction was used in a PCR reaction using degenerate sense and antisense oligonucleotides (1 μ M each), 25 μ M of each nucleotide and 2.5 units Taq polymerase (Boehringer). 30 cycles were performed with 1 min for each 94°C, 50°C and 72°C step. Fragments of approximately 250 bp were gel purified, cloned in Bluescript and sequenced.

mCLK2, mCLK3 and mCLK4 were cloned from a mouse embryo 11.5 p.c. 1ZAP cDNA library (Ciossek et al., 1995, Oncogene 11:2085-2095) using the isolated PCR fragment as a probe according to manufacturer's instructions (final wash in 0.5xSSC/0.1%SDS at 42°C) (Stratagene). mCLK1 was cloned by reverse transcriptase PCR from 1 μ g brain poly (A)⁺ RNA using specific primers mCLK1s-Bam, CGGGATCCCTTCGCCTTGCAGCTTTGTC (SEQ ID NO. 3) and mCLK1as-EcoRI, CGGAATTCCTAGACTGATACAGTCTGTAAG (SEQ ID NO. 4), and Pwo polymerase (Boehringer).

DNA sequencing was performed using the dideoxynucleotide chain termination method (Sanger et al., 1977, *Proc. Natl. Acad. Sci. USA* 74:5463-5467) using sequenase, reagents and protocols supplied by United States Biochemical Corporation. Comparisons of the deduced protein sequences were carried out using MacDNASIS PRO (Hitachi) software. Amino acid alignments were constructed using a Waterman algorithm.

EXAMPLE 2: TISSUE DISTRIBUTION OF CLK SERINE/THREONINE KINASES

Expression patterns of CLK kinase genes (including the previously cloned mCLK1) were analyzed by Northern blot hybridization of total RNA from selected mouse tissues, as well as from different mouse tumor cell lines. Each CLK gene was detected in all investigated tissues, although the expression patterns were different for each gene. The expected size of the full-length mRNA is ~1.8kb for all CLK kinases, and this was detected in all tissues and cell lines, albeit at different levels.

A doublet was detected at around 1.8kb, whereby the upper band represents the message of the full length protein and the lower one is likely to be the alternatively spliced form, responsible for a truncated, catalytically inactive protein. Duncan et al., 1995, *J. Biol. Chem.* 270:21524-21531. Commensurate with this alternative splice, differences in the ratio of the two alternatively spliced messages were detected for each CLK gene.

Differences in expression patterns were observed for the CLK genes, especially in testes. Low mCLK1 expression levels were observed in testes as compared to mCLK2, mCLK3 and mCLK4. However, while almost all of the mCLK3 message represented the catalytically active splice form, mCLK4 was expressed predominantly as a message encoding the truncated protein. mCLK2 was also highly expressed in this tissue, but as a larger transcript. Similar large

transcripts, which did not correspond to the expected message size, were detected for all mCLK genes which most likely represented non- or partially spliced messages in analogy to mCLK1. Duncan et al., 1995, *J. Biol. Chem.* 270:21524-21531. The ratio of these larger RNA species, when compared to the coding mRNA, varied among the CLK kinases.

Because it was reported (Ben-David et al., 1991, *EMBO J.* 10:317-325) that mCLK1 kinase was over-expressed in certain cancer cell lines, studies were extended to mCLK1-4. Although messages for the four genes were detected in all cell lines tested, albeit in sometimes very low quantities, significant differences of expression levels between the cell lines for each individual gene were observed. However, an overall increase of mCLK mRNA was not detected in transformed cells, even though higher levels of particular mCLK messages were detected in some cell. Low expression levels were detected in Hybridoma, WEHI and NF561 cell lines, with the majority of the messages representing the splice form encoding the truncated product. The mRNA expression levels of mCLK1-4 genes were investigated in the C2C12 cell line and Li adipocytes during differentiation, but no noticeable change in expression was detected.

RNA was extracted from frozen adult mice tissues or tissue culture cells. Puissant and Houdebine, 1990, *Biotechniques* 8:148-149. 10µg total RNA was then electrophoresed in 1.2% agarose formaldehyde gels

(Sambrook et al., 1989, *Cold Spring Harbour Laboratory Press*) and transferred to Hybond N membranes (Amersham). Hybridization was performed overnight in 50% formamide, 5x SSC (750mM sodium chloride, 75mM sodium citrate), 5x Denhardt's (0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, 0.1%BSA), 0.2% SDS and 100µg/ml salmon sperm DNA. $1-3 \times 10^6$ cpm /ml of ^{32}p -random primed DNA probe (Amersham Megaprime kit) was used, followed by washes at 0.2xSSC/0.1%SDS at 42°C. Blots were incubated with Hyperfilm-MP (Amersham) at -80°C for 2 weeks. Membranes were stripped for reuse by boiling in 0.1% SDS/water.

EXAMPLE 3: EXPRESSION OF FUNCTIONALLY ACTIVE CLK PROTEIN KINASES

Glutathione S-transferase (GST) mCLK1-4 fusion constructs were generated to investigate the catalytic activity of these protein kinases. These protein kinases were cloned from pcDNA and expressed *in vitro*. The expression levels were almost identical and full-length fusion proteins of the expected molecular weights were obtained. The GST fusion proteins were purified on glutathione-sepharose beads and utilized to perform *in vitro* kinase assays using myelin basic protein or histone H1 as substrates. All constructs were catalytically active, autophosphorylated, and the levels of activity were clearly above the background seen from an equivalent amount of *in vitro* produced GST protein alone.

Catalytically inactive lysine to arginine mutants could not phosphorylate any substrate above background phosphorylation. However, mCLK1 and mCLK4 displayed a dramatic difference in enzymatic activity versus mCLK2 and mCLK3. This observation was consistent even when changing a variety of buffer conditions during fusion protein purification and assays or when changing metal ion concentrations. Several fold changes in kinase activity were observed due to the conditions used, but differences in enzymatic activities seen between the two groups of mCLK kinases persisted.

Phosphorylation specificity of mCLK1-4 protein kinases were examined and compared using biochemically purified and dephosphorylated SR proteins as substrates. SR proteins were purified from 5×10^9 log phase F-MEL suspension cells according to standard procedures. An aliquot of the purified proteins was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to confirm that the expected proteins were purified to near homogeneity.

Following dephosphorylation by the protein phosphatase 1 γ catalytic subunit, SR proteins were used as substrates for the *in vitro* produced and purified GST-mCLK1-4 fusion proteins in an *in vitro* kinase assay. All mCLK kinases were able to phosphorylate SRp20, SRp30a and to a lesser extent SRp40 and SRp55. The lower signal of SRp40 and SRp55 relative to SRp20 and SRp30 most likely reflected the lower quantity of these proteins.

SRp75 was not visualized in these experiments since the autophosphorylated mCLK proteins migrated at the same position. mCLK1 and mCLK4 phosphorylated SRp30a (upper band) more strongly than SRp30b, whereas mCLK2 and mCLK3 phosphorylated both with almost equal efficiency. A marked difference in catalytic activity was visualized between mCLK1 and mCLK4 versus mCLK2 and mCLK3, despite equal amounts of protein in each assay.

To investigate the specificity of mCLK kinases, recombinant human peptidyl-prolyl isomerase PIN1 was utilized as a substrate. Lu et al., 1996, *Nature* 380:544-547. Although it also contains several amino terminal serine and arginine residues and is localized to nuclear speckles, neither of the mCLK kinases was able to phosphorylate this protein *in vitro*.

GST fusion constructs were generated by subcloning full length mCLK1, mCLK2, mCLK3 and mCLK4 cDNAs by PCR into pGEX vectors (Pharmacia), creating in-frame glutathione S-transferase (GST) fusion constructs using the following primers for PCR: mCLK1s-Bam (as above); mCLK1as-Not I, TATAGCGGCCGCTAGACTGATACAGTCTGT (SEQ ID NO. 5); mCLK2s-Sma I, TCCCCCGGGATGCCCCATCCCCGAAGGTACCA (SEQ ID NO. 6); mCLK2as-Not I, TATAGCGGCCGCTACCGACTGATATCCCGACTGGAGTC (SEQ ID NO. 7); mCLK3s-Sma I, TCCCCCGGGGAGACGATGCATCACTGTAAG (SEQ ID NO. 8); mCLK3as-Not I, TATAGCGGCCGCGCTGGCCTGCACCTGTCATCTGCTGGG (SEQ ID NO. 9);

mCLK4s-EcoRI, CGGAATTCATGCGGCATTCCAAACGAACTC (SEQ ID NO. 10), mCLK4as-Not I, TATAGCGGCCGCCCTGACTCCCACTCATTTCTTTTAA (SEQ ID NO. 11). The cDNAs encoding the fusion construct were then recloned in pcDNA3 (Invitrogen) by PCR using the GST upstream primers: GST-EcoRI, - CGGAATTCGCCACCATGGCCCCTATACTAGGTTAT (SEQ ID NO. 12) (for mCLK1) and GST-Hind III, GCCAAGCTTGCCACCATGGCCCCTATACTAGGTTAT (SEQ ID NO. 13) (for mCLK2, mCLK3 and mCLK4).

Integrity of the clones was checked by sequencing and by a coupled transcription-translation assay using T7 RNA polymerase and rabbit reticulocyte lysate according to the manufacturer's protocol (Promega).

mCLK 1-4 mutants containing a lysine (K) to arginine (R) substitution at position 190 (mCLK1), 192 (mCLK2), 186 (mCLK3) and 189 (mCLK4) were generated using a site-directed mutagenesis protocol. Kunkel, 1985, *Proc. Natl. Acad. Sci. USA* 82:488-492. Oligonucleotide primers were as follows: (mCLK1-K190R) GTAGCAGTAAGAATAGTTAAA; (mCLK2-K192R) (SEQ ID NO. 14) GTTGCCCTGAGGATCATTAAGAAT; (mCLK3-K186R) (SEQ ID NO. 15) GTTGCCCTGAGGATCATCCGGAAT; (mCLK4-K189R) (SEQ ID NO. 16) TACAATTCTCACTGCTACATGTAAGCCATC (SEQ ID NO. 17).

³⁵S-methionine labeled GST-mCLK1-4 fusion proteins were produced in a 50gl coupled *in vitro* transcription/translation reaction using manufacturer's instructions (Promega).

2g1 of each lysate was checked and quantified for the amounts of produced protein by SDS-PAGE and autoradiography. Equal amounts (usually 20-30, μ l of lysate) were added to 500 μ l PBS (1mM PMSF, 10 μ g/ml aprotinine), 30 μ l of GSH-sepharose beads (Pharmacia) and incubated on a rotating wheel for 2 hours at 4°C. This step resulted in quantitative binding of the fusion proteins. The beads were then washed three times in 500 μ l PBS and once in 500 μ l kinase assay buffer (20mM Hepes, 10mM $MgCl_2$, 1mM DTT, 200 μ M sodium orthovanadate, 1mM EGTA, pH 7.5). The assay was carried out for 30 minutes at room temperature in 30 μ l kinase assay buffer with 20 μ M ATP, 4 μ Ci γ - 32 P-ATP (Amersham, 10mCi/ml) and 14.5 μ g of myelin basic protein (Sigma) or histone H1 (Boehringer) respectively.

SR protein kinase assays were essentially carried out as described above, except that ~2.5 μ g of dephosphorylated SR proteins were used and that the kinase assay buffer also contained 1 μ M Microcystin LR (Sigma). The reaction was stopped by adding 30 μ l of 2xSDS sample buffer. The samples were boiled for 5 minutes and 15 μ l were loaded on a 15% SDS-polyacrylamide gel. Following electrophoresis, the gels were stained with Coomassie, dried and exposed to Hyperfilm -MP (Amersham) for 24 hours. The 35 S-methionine signal was suppressed with a 3M Whatman paper placed between the film and the gel.

SR proteins were purified from 5×10^9 Friend murine erythroleukemia cells (F-MEL) according to the protocol

described (Zahler et al., 1992, *Genes Dev* 6:837-847) and resuspended in buffer D. Dignam et al., 1983, *Nucleic Acids Res.* 11:1475-1489. 30 μ l of SR proteins (~0.5 μ g/ μ l) were incubated on ice for 10 minutes in 0.7mM MnCl₂ and 5mU Protein Phosphatase 1 γ -catalytic subunit (Boehringer), followed by 60 minutes at 30°C. Mermoud et al., 1994, *EMBO J.* 13:5679-5688. 5 μ l of dephosphorylated SR proteins were used per assay.

EXAMPLE 4: NUCLEIC ACID PROBES, METHODS, AND KITS FOR DETECTION OF CLK KINASES AND OTHER RELATED POLYPEPTIDES

A nucleic acid probe of the present invention may be used to probe an appropriate chromosomal or cDNA library by usual hybridization methods to obtain other nucleic acid molecules of the present invention. A chromosomal DNA or cDNA library may be prepared from appropriate cells according to recognized methods in the art (cf. "Molecular Cloning: A Laboratory Manual", second edition, Cold Spring Harbor Laboratory, Sambrook, Fritsch, & Maniatis, eds., 1989).

In the alternative, chemical synthesis can be carried out in order to obtain nucleic acid probes having nucleotide sequences which correspond to N-terminal and C-terminal portions of the amino acid sequence of the polypeptide of interest. The synthesized nucleic acid probes may be used as primers in a polymerase chain reaction (PCR) carried out in accordance with recognized

PCR techniques, essentially according to PCR Protocols, "A Guide to Methods and Applications", Academic Press, Michael, et al., eds., 1990, utilizing the appropriate chromosomal or cDNA library to obtain the fragment of the present invention.

One skilled in the art can readily design such probes based on the sequence disclosed herein using methods of computer alignment and sequence analysis known in the art ("Molecular Cloning: A Laboratory Manual", 1989, supra). The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes may be visualized using known methods.

The nucleic acid probes of the present invention include RNA, as well as DNA probes, such probes being generated using techniques known in the art. The nucleic acid probe may be immobilized on a solid support. Examples of such solid supports include, but are not limited to, plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, and acrylic resins, such as polyacrylamide and latex beads. Techniques for coupling nucleic acid probes to such solid supports are well known in the art.

The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample that is compatible with the method utilized.

One method of detecting the presence of nucleic acids of the invention in a sample comprises (a) contacting said sample with the above-described nucleic acid probe under conditions such that hybridization occurs, and (b) detecting the presence of said probe bound to said nucleic acid molecule. One skilled in the art would select the nucleic acid probe according to techniques known in the art as described above. Samples to be tested include but should not be limited to RNA samples of human tissue.

A kit for detecting the presence of nucleic acids of the invention in a sample comprises at least one container means having disposed therein the above-described nucleic acid probe. The kit may further comprise other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound nucleic acid probe. Examples of detection reagents

include, but are not limited to radiolabelled probes, enzymatic labeled probes (horseradish peroxidase, alkaline phosphatase), and affinity labeled probes (biotin, avidin, or streptavidin). Preferably, the kit further comprises instructions for use.

A compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers include a container that accepts the test sample, a container that contains the probe or primers used in the assay, containers that contain wash reagents (such as phosphate buffered saline, Tris-buffers, and the like), and containers that contain the reagents used to detect the hybridized probe, bound antibody, amplified product, or the like. One skilled in the art will readily recognize that the nucleic acid probes described in the present invention can readily be incorporated into one of the established kit formats that are well known in the art.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The molecular complexes and the methods, procedures, treatments, molecules, specific compounds described herein are presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising", "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description

and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed.

In particular, although some formulations described herein have been identified by the excipients added to the formulations, the invention is meant to also cover the final formulation formed by the combination of these excipients. Specifically, the invention includes formulations in which one to all of the added excipients undergo a reaction during formulation and are no longer present in the final formulation, or are present in modified forms.

In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine, chlorine, and iodine, claims for X being bromine and claims for X being bromine and chlorine are fully described.

Other embodiments are encompassed by the following claims.